DOCUMENT RESUME

ED 331 870 TM 016 413

AUTHOR Fisher, William P., Jr.
TITLE Conversing, Testing, Questioning.
PUB DATE Mar 90
NOTE 38p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, March 16-20, 1990).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; Learning Processes; *Measurement Techniques; *Models; *Questioning Techniques; Teaching Methods; *Test Use
IDENTIFIERS Conversation Theory; Rasch Model

ABSTRACT

Educational measurement is presented as integral to the process of entering into and guiding conversations in which teachers and students participate. A measurement model patterned on the playful, interactive back-and-forth motion of conversation could aid in overcoming problems in teacher-student interactions. According to H. G. Gadamer, conversation is an art of testing because the partners to a dialogue must establish common reference points and standards of comparison in order to communicate. Testing is an inherently circular and two-sided process; to know which questions are appropriate, the examiner has to have asked some questions. Educational measurement becomes conversational when it makes the articulation of the object of discourse and the negotiation of meaning the first task of measurement. The measurement models of G. Rasch and B. D. Wright allow the object of discourse to represent itself through the players involved. Measurement in this sense can be an important part of conversation between teachers and students. An 85-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331870

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

WILLIAM P. FISHER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Conversing, Testing, Questioning

William P. Fisher, Jr.
Marianjoy Rehabilitation Center
March 1990

Paper accepted for presentation to Division G of the AERA during session 53.43 at the 1990 Annual Meeting in Boston.

TM016413

BEST COPY AVAILABLE

This essay was not originally intended to address the problem of dropping out. The placement of the paper in this session led to some extra effort aimed at relating its content to the overall theme. The reader must keep in mind that the author does not have a specialized interest in dropouts or at-risk students, but has made a small effort toward bringing what originated as a philosophical and methodological analysis to bear on the problem at hand.

Introduction. The purpose of this paper is to present educational measurement as an important part of the process of entering into and guiding the conversations in which teachers and students participate. Educational measurement is most often practiced by means of tests made up of questions to which students respond; this exchange of question and answer is a rudimentary model of conversational interaction that can be used to expand and improve the participation of students in educational activities. Finn's (1989) model of dropping out as a developmental process beginning in the earliest grades suggests that there are four levels of participation in school activities; at the second level, "students initiate questions and dialogue with the teacher and display enthusiasm ..." (Finn 1989, p. 128). Philosophically speaking, each of the other three levels of participation (passive acquiescence, extracurricular activity, and roles in governance) are manifestations of more or less questioning of, and dialogue with, the powers that be. The problem is that there are so many participants to these dialogues, and that there are so many dialogues, each often occurring on several levels simultaneously, that it can be very difficult to keep track of what has been said, who said it, whose turn it is to speak or listen, who has something important to say, or even what the object of discourse is. This essay will attempt to show that a measurement model patterned on the playful, interactive back-and-forth motion of conversation could be a great aid in overcoming these problems. The focus will be on showing that conversation is itself an art of testing, and that this sense of testing can be used as a model of understanding applicable to the goals of educational measurement.

The first obstacle that a conversational model of educational measurement is likely to encounter concerns the apparently one-sided nature of the question and answer process in testing, as opposed to the two-sided nature of question and answer in conversation. The power relations that frame the dialogues within the school are themselves shaped by relations in the larger social, political and economic context, making for what frequently appear to be grossly imbalanced, if not completely one-sided, interactions. Recent researches in other fields, however, such as in the ethnography of post-colonial societies (Comaroff 1985), have shown that resistance to authority does not always follow orthodox revolutionary strategy. In the same way, administrators, teachers and students devise strategies for coping with situations too large for them to control directly (Woods 1989). Similarly, the traditional conception of education as a one-way flow of information from the master to the novice "tends to induce researchers to underestimate the two-way processes that regulate communication between teacher and learner" (Perret-Clermont and Schubauer-Leoni 1989, p. 579). More and more researchers are recognizing that learning is a negotiation of meaning (Carraher 1989, Woods 1989, Voigt 1989), "an articulation of the object of discourse" (Schubauer-Leoni, et al 1989), and that its direction is ruled from within the educational dialogue itself, not from without by an authority external to the interchange.

Though it has yet to have much of an effect on the way educational measurement is understood, the focus on the emergence of meaning and on discourse processes has also influenced the way philosophers and historians think about measurement and technology (Heidegger 1977b, Kuhn 1961, Heelan 1983, 1985, Ihde 1979, Ackermann 1985, Andrich 1988, Duncan 1984, Ormiston and Sassower 1989). According to this school of thought, measurement is most fundamentally an effort aimed at enframing the sphere in which phenomena are created (Hacking 1983, Heelan 1989), by checking and re-checking that the object of conversation remains relatively and probabilistically constant across laboratories, researchers, linguistic barriers, treatment samples, instruments, etc. This constancy is true and objective only in the sense that

it provides a likely story, a fictive account, of what is a reasonable thing to think about what happened in the question and answer exchange. Such a story is an expression of the line of inquiry delineated and plotted by the way the object plays itself out across specific contexts within a larger context.

In the same way that educators need to overcome the notion that education is a one-way process of information transfer and control, philosophy has come to terms with the positivist notion that humans are irrevocably stuck in a one-way relation to facts. Instead of being mere recipients of what is eternally given, scientists are engaged in a dialogue with nature, negotiating meaning among themselves by entertaining various models and metaphors as potentially apt articulations of the way phenomenological objects come into being (Toulmin 1982, Livingston 1986, Holton 1988, Latour and Woolgar 1979; Black 1962). This dialogue is a struggle to force nature to respond to the questions scientists put to her, to compel her to poetically suggest some metaphor by which the arrow of meaning can be delineated; "but that is not to say either that nature will respond to any theory at all or that she will ever respond very much" (Kuhn 1961, p. 176; also see Kuhn 1979, Boyd 1979, Rothbart 1984, Gerhart and Russell 1984).

Many of Kuhn's critics have become bogged down by the idea that nature responds in any way to the theoretical predispositions of scientists; to them the language of metaphors and fictions seems to be a complete relinquishment of any hope for objectivity. But objectivity does not have to be constituted in a one-sided fashion. More than anything else, science is a process of articulating new objects of discourse in the language of what is already known; philosophy has posed the question of how that articulation comes about since its origin in ancient Greece. Husserl (1970) analyzed this problem in terms of Galileo's "fateful omission" of the means by which the modern science of measurement was formulated. Husserl's student, Heidegger, put it in terms of the "question of Being:" where Descartes (1971, p. 184) was willing to leave the "simple," but necessarily prior, cognizance of what counts as knowledge and existence off the

record, Heidegger's (1982, p. 125) work is an effort aimed at putting this back on the record. In the same vein, Heidegger's student, Gadamer (1989), sought to counter the Enlightenment prejudice against prejudice; from this follows his comment that "when Thomas Kuhn published *The Structure of Scientific Revolutions*, I was very pleased because it supported my view" (Gadamer 1975, p. 336).

Construing measurement as an extension of conversation has great potential for positively influencing educational measurement by helping us keep track of what the object of discourse is, and where the participants in the dialogue are positioned in relation to it. As the sociologist and measurement historian O. D. Duncan (1984, p. 221) has pointed out, "the social roots of social measurement are in the social process itself," meaning that the social relations through which meanings are negotiated and objects of discourse are articulated are the surest clues we have as to what is most important to the conceptualization of measurement. When the plant (social measurement) is separated from its roots (the social process), it will wither (Duncan 1984, p. 237), as it long since has; conversely, this separation also has an adverse effect upon the social process roots, as the phenomenon of dropping out attests. Measurement is too often one-sided and manipulative because our understanding of it detaches it from its source in social life, and then social life suffers because our measuring processes are incapable of adequately articulating objects of discourse, of analyzing them and of monitoring the extent to which we turn a listening ear and give a clear voice to them.

So the question is, how can testing be made more conversational? What does it mean for measurement to be the articulation of objects of discourse, to aid in the negotiation of meaning, and for it to be rooted in social processes? These questions are asked as preliminary orientations toward research that would test whether identification with school and extensive involvement in school activities are actually adequate indicators of the propensity to drop out. These questions also apply to the way the research itself is carried out because social measurement in general does

little to attend to the object of the question and answer process, as Duncan (1984) reminds us. This line of questioning will be pursued through an examination of how conversation is itself an art of testing. Then, the ways in which testing and measurement can be conversationally construed will be explored, and contrasted with non-dialogical positions. Finally, critics who contend that the conversational approach is not sufficiently radical will be addressed, and the conversational model will be defended and qualified in light of these criticisms.

Conversing as testing. "Conversation ... is an art of testing. But the art of testing is an art of questioning." This statement appears in Hans-Georg Gadamer's (1989, p. 367) work Truth and Method in the context of his explication of interpretation theory. Gadamer takes conversation to be an art of testing because the partners to a dialogue must establish common reference points and standards of comparison in order to communicate.

> To conduct a conversation requires first of all that the partners to it do not talk at cross purposes. Hence its necessary structure is that of question and answer. The first condition of the art of conversation is to ensure that the other person is with us. We know this only too well from the reiterated yesses of the Platonic dialogues. The positive side of this monotony is the inner logic with which the development of the topic proceeds in conversation (Gadamer 1989, p. 367).

A topic develops in conversation insofar as each locution is checked for its reference to the object of inquiry; when the limits of, and discrepancies within, references arise, so do new questions. A series of questions that all more or less attend to the same issue unfolds into a line of inquiry, which provides the interlocutors with opportunities to keep track of what has been said and of each other. The path followed in a question and answer exchange is ruled, delineated and dominated only by the self-representative, playful activity of the thing itself. Its logic is monotonous in the same way that walking is monotonous; one step follows another of its own accord, but there is more than mere repetition going on. Walking and talking are recreative

activities in which participants are transformed to the extent that they give themselves up to the activity of the thing itself.

We know what the topic of conversation is, and where the speakers are in relation to it, only by means of the fictive account rendered by each step in the exchange of questions and answers. Someone once described literary fictions as accounts that are true even if they did not happen, which is to say that stories help us to understand who we are not because the specific things said actually happened but because their general pattern describes an experience that people share. In the same way, communication and community ha[illegible]en not because we have absolute grounding in eternal facts to which we must submit, but because of the way that myths, metaphors and models provide a linguistic tension that holds us in our social orbits. A model of educational measurement that could provide such stories has the potential of helping to alleviate problems concerning students' identification with, and participation in, schools.

Gadamer takes conversing as an art of testing in two ways. First, the topic of conversation is tested for viability through a reiterated elaboration of its implications within individual dialogues and across them. An idea may be shown to be a dead end by the questions, or lack of them, to which it gives rise within a single conversation, or it may play itself out only through extensive applications. In the conversation that artfully tests what is said for adherence to the object, questions and answers stick to the topic. By constantly referring and deferring to the object dominating the interaction, what is said unfolds the inner structure and logic of the topic. As the line of inquiry develops, the interlocutors turn together under its influence in the same social orbit.

Second, the partners to dialogue also need to keep track of each other. They are therefore tested for their abilities to articulate and respond to the questions arising from the topic. Respondents inexperienced with an idea may prematurely reject or accept its viability because they are unable to perceive and pursue its implications; this may be the most obvious sense in

which conversation is a test -- the ability of the respondent to correctly answer the questions put is at issue. However, the questioner is also tested for the capacity to stick to the topic, and to be perceptive enough to know what question to ask next, even when that question endangers his or her own opinion.

Gadamer (1989, p. 363; 1976, pp. 12-13) points out that Socrates' partners in conversation often felt that his role as questioner gave him an unfair advantage by allowing him to choose the questions. They also figured it would be easier to ask than answer because there would be less risk of displaying ignorance. However, Plato took delight in showing how fast a dialogue could come to a halt when the questioner lacks imagination or is pursuing a hidden agenda, and is therefore unable to let the object of discourse dominate the interchange. A conversation is not a test when interlocutors are unable to keep in touch with one another or if they speak at cross purposes. Partners to a conversation are most likely to have one of these things happen to them if the questions posed are not framed in the interest of creating meaning, but are instead pushing a hidden agenda such that an openness to new possibilities is closed off.

Slanted questions are so-called because they only seem to lead toward the object of inquiry. Conversations are tests when they establish the viability of an idea's arrow of meaning by exploring how far it is possible to follow and clear the path of its argument. As is made so plain by the double meaning of the French word for meaning, *sens*, "sense involves a sense of direction. Hence the sense of the question is the only direction from which the answer can be given if it is to make sense" (Gadamer 1989, p. 362). Viability takes the form of a general invariance to the order of the questions in the line of inquiry; questions that do not point together toward an open issue are slanted in a direction away from the actual question that remains implied.

So we see that conversation is a test primarily in the sense that what is said is tested for its viability; in the Socratic sense of midwifery, the teacher's role is to aid in the birth of ideas, and to hold them up to the light and inspect them to see that they are well-formed and able to play

themselves out for whatever they are worth. Conversation is a test of a person's ability to respond to questions only insofar as it is also a test of the difficulty of the questions because it is what is said -- the questions and the answers -- that must first be tested for adherence to the object if ability and difficulty are to mean anything. If what is said does not lend itself to organization as a likely account in which signs hang together, meaning remains out of reach. When a series of questions asked of different persons results in sets of reiterated responses that share a common monotony, these responses represent the inner logic by which the articulation of the object of discourse has proceeded. Within a culture, things, as the objects of conversations, must maintain a relative and probable invariance in the directions they take up for them to exist as such. It is in the determination of the extent to which the partners to a conversation are talking about the same thing that the art of testing finds its pla e.

When questions and answers share a common unit of meaning that repeats itself at regular intervals along the dimension delineated, it is said that a line of thought has been drawn out, that something has been gotten straight and that things add up. What we need in order to transform these metaphors of virtual measurement into a structure for actual measurement are ways 1) of establishing the convergence of questions and answers upon a single line of thought, and 2) of checking to see that the topic remains relatively and probabilistically invariant within and across dialogues. When questions and answers converge upon a common object such that an arrow of meaning is delineated, they all point in the same direction. When we further see that similar, but not identical, questions can be asked by other people in other similarly constituted situations with approximately the same results, it can be said that a single object of discourse dominates each of the different dialogues. The constancy of meaning that holds across minor contextual changes in time, place, persons and articulations of an object of discourse is the condition of the possibility of language and culture. The convergence of question and answer, accompanied by the separation of the meaning of the questions asked from the particulars of their representation and

from the particular persons responding, is referred to in measurement theory as sample-free instrument calibration and instrument-free person measurement (in short, sample-free measurement) (Wright 1968, 1977, 1984, 1985). The purpose of the next section is to explicate a two-sided, conversational, and fictive construal of sample-free measurement.

Testing as conversing. The critics of Kuhn and Heidegger fear the loss of all objective standards if some way out of the interactive dynamic unity of subject and object is not found. However, the fact that all learning takes place through what is already known is one of the most fundamental Platonic motifs and is key to all academic understanding. The application of the Socratic method requires that the object of inquiry be allowed to unfold such that what the student already knows is transformed into something new. The same sort of thing happens in what is known as the targeting of a test: making sure that the questions are neither too easy nor too hard for the students, but are aimed directly at their level of ability in such a way that they are challenged, but still are confident of their ability to respond successfully to at least some of the questions.

In order to see that everyone is talking about the same thing, the questions asked must be united according to a theory of what each of them is supposed to contribute toward the delineation of a path of meaning, how it will lead toward the articulation of an object of discourse, and be a step in the negotiation of meaning. The theory is a fiction that guides the focusing of questions and the organization of data; it follows from the fictive quality of theory that "no experiment gives quite the expected numerical result," a fact that is jokingly referred to as the Fifth Law of Thermodynamics (Kuhn 1961, p. 165). When data conform too closely to theory, suspicions regarding the veracity of the data, or the non-triviality of the theory, are justified. As Rasch (1960, p. 115) points out, even the predictions made by deterministic measurement models are not absolutely true, but are fictional truths that actual data only approximate, falling into a probability distribution around the predicted value. The fact is that "a model is not meant to be

things are always understood in terms of something already known, which is the reason for Gadamer's (1989, p. 362-379) emphasis on the hermeneutic priority of the question. There is always an implicit question that resides in each and every recognition of things for what they are; to recognize something for what it is is to be able to answer the question, "What is that?" The art of testing is an art of questioning because of the ubiquitous, though usually implicit, presence of questions in every application of the fore-structures of knowledge (Gadamer 1989, p. 367). Science originated in the form of questioning that is willing to leave open the possibility of the confirmation or refutation of the assertion that something *is* one way or another; Gadamer (1989) points at the important role that the openness of questioning to possible refutation has played in the philosophy of science from the work of Francis Bacon to that of Karl Popper. Questioning is the fundamentally Socratic aspect of science, the willingness to expose one's opinions as interpretations to the critical test of the way the thing itself playfully rules over and leads the question and answer process along the line of inquiry.

Educational tests are too rarely used in a manner that suggests that the questioner and the respondents hold a common interest in the articulation of the object of discourse. It is generally assumed that the questions asked are appropriate and that they in fact stick to the subject. Heidegger (1962, p. 195), however, reminds us that

> our first, last and constant task is never to allow our fore-having, fore-sight, and fore-conception to be presented to us by fancies and popular conceptions, but rather to make the scientific theme secure by working out these fore-structures in terms of the things themselves.

Gadamer's sense of conversation as an art of testing is an exposition of what Heidegger called the hermeneutics of facticity - the manner in which the fore-structures of knowledge are worked out in terms of the things themselves. I have no data that show dropping out to be a function of the fact that educational measurement only rarely works out the fore-structures of thought in terms of the things themselves, but is not it highly likely that many of our measurement problems

true" (Rasch 1964: 24, 2, 3; 1960: 37-38). Models are meant to bear meaning, to be heuristic fictions redescribing reality and guiding the conversations between researcher and researched, and among researchers (Hesse 1970, Black 1962, Ricoeur 1977: 239-246, 1981: 185-190).

The probabilistic truth associated with mathematical models is thus aptly called a fiction that has the heuristic value of guiding inquiry, of pointing a direction in which questioning may be productive of new insights. This guidance focuses the intention of the research by determining in advance what kind of questions will define the variable and provide a measure. Each respondent's reiterated answers develop the line of questioning to the extent that they share a monotonous inner logic, expressed as an order through which the pertinence and difficulty of each question is evaluated just as much as the pertinence and ability of the student are. When the questions and responses share a common order, an invariantly structured arrow of meaning has been delineated into a test instrumental to the measurement of ability. To be instrumental in this way is to have allowed the unity of question and answer to unfold according to its own emergent possibilities, taking up the direction it does not because the questioner's preconceptions present themselves as the object of discourse, but because preconceptions inevitably exist in a prior unity with the thing itself. By allowing preconceptions to hold sway, the object is articulated; then the meaning of the preconceptions and other unsaid contextual nuances can be negotiated by checking to see if the questions and answers point together along a common path of meaning.

Therefore, contrary to the perception of testing as one-sided, we see that it is inherently circular and two-sided: in order to ask test questions, the examiner must have some prior idea of what questions are appropriate, and in order to know which questions are appropriate, the examiner has to have asked some questions. This is why Rasch (1960, p. 110) did not "find it feasible to introduce the two concepts [of ability and difficulty] separately, [but] had to formalize them simultaneously ... one by means of the other, as it were, and vice versa, without getting into any logical circle." The non-vicious quality of the hermeneutic circle follows from the fact that

result from the way we allow fancies and popular conceptions to pass for standards of comparison, and that the resulting double and triple and quadruple standards imposed upon those measured foster more alienation than they do identification?

The history of science shows that a great deal of qualitative analysis of phenomena is accomplished before quantitative measurement becomes possible; is it not likely that the current interest in qualitative explications of the contextual factors involved in the way things are is a positive and necessary step toward obtaining the detailed understandings required for scientific theorizing and analysis? Interpretive and qualitative methods follow "the hermeneutical rule that we must understand the whole in terms of the detail and the detail in terms of the whole" (Gadamer 1989: 291), just as the question and answer process in conversation must. According to this rule, understanding depends upon the cultural and linguistic preconceptions and assumptions that are brought to bear on each part of it. Perception, observation and comprehension are always a matter of seeing things *as* something, in the terms of a particular historical, cultural and linguistic context.

It is only within this general context that factuality becomes possible. In order for an understanding or observation to be communicated it must be repeatable, and each replication of it constitutes a test of its veracity. The basic difference between the hermeneutic and positivist accounts of science is that the former does not claim absolute generality to facts, but recognizes the larger context of their origin and application. There is nonetheless a sense in which facts are not completely context-dependent; this was analyzed by Heidegger in terms of the hermeneutics of facticity, which is what Gadamer is implicitly referring to in his sense of conversation as an art of testing. In order to be seen at all, things are seen as something, in terms of a conceptual structure that is not the thing *in* itself, and which changes, sometimes slowly and sometimes rapidly, from one historical period or culture to another. The hermeneutic point is that the rules of language are shaped just as much as they are followed by usage. The object of the continuing

conversation changes as it unfolds, transforming itself as well as those who allow themselves to be conducted by it. The transformations involved in the negotiation of meaning may lead gradually or suddenly to the need for new models of understanding, but lead on they will, and this is one of the strongest arguments in favor of modeling measurement in such a way that it allows room for being transformed by the internal critiques that follow from the chance and accidental events normally found in human behavior. We stand in need of a way to constantly monitor the objects of our ongoing conversations; a probabilistic model of simultaneous conjoint measurement meets this need.

When is educational measurement conversational, and when is it not? The short answer to this question is that educational measurement is conversational when it makes the articulation of the object of discourse and the negotiation of its meaning the first task of measurement. Only when it can be ascertained that the questions asked and the persons responding all speak to a single object can the secondary goal of comparing the difficulty of items and the abilities of persons be properly begun. More often than not, however, practitioners of educational measurement one-sidedly presume that tests and curricula stick to the topic in question. Test questions are usually held to be unquestionably authoritative and legitimate because forces outside of the school district, the school and the classroom -- such as public opinion, legislative mandates, textbook publishing practices and the authority of testing agencies -- have represented them as such on the basis of political and economic reasons that are supported by an allegedly scientific thinking to which few of those affected have access. Similarly, teachers too rarely question the authority of the questions they write; this may be because they have not been trained to do otherwise, for one thing, but it is also because it seems reasonable for teachers to assume that they know more about the subject than the students do.

The point of the preceding pages, however, has been that this attitude prevents teachers from fully engaging themselves with the students in the study and articulation of the object of discourse. The situation is analogous to that of the debate between Socrates and the Sophists; teachers commonly practice sophistry by conducting a one-sided monologue in which communication and learning are presumed to be effected by a one-way flow of information, an inscription of knowledge upon the impressionable blank slates of the students' minds that can then be tested by requiring the students to reverse the flow of information. Rasch and Wright's theory of a simultaneous and conjoint fundamental measurement procedure, on the other hand, offers the more Socratic option of checking the extent to which what has been said, done and written (the data) lives up to its billing as representative of the thing itself, as a sign that the thing everyone presumes to have dominated the conversation actually did so. This strategy aims at the articulation of the thing itself as a primary concern that precedes the comparison of measures.

Educational measurement realizes its innermost possibilities for the creation of meaning, when it embodies the structure of conversation as a test:

> To conduct a conversation means to allow oneself to be conducted by the subject matter to which the partners in the dialogue are oriented. It requires that one does not try to argue the other person down but that one really considers the weight of the other's opinion. Hence it is an art of testing (Gadamer 1989, p. 367).

In the current state of testing, questioners too frequently argue the respondents down, with no systematic effort put toward considering the weight of what has been said. Teachers and tests are often allowed to un-Socratically presume to be able to lead students to knowledge, to be able to conduct a conversation over and against the way in which the subject matter itself dominates the question and answer exchange. In such a context, the weight of the students' opinions is considered only in the roundabout manner through which learning is 'scientifically' studied, debated, and influenced by extraneous political intrigues, ideological programs and economic interests before coming back to the students in the form of yet another externally imposed

structural force. The point is not that such interests can be eliminated or avoided, but that the authentic voice of authority is that one which emerges from within the local, intrinsic and indigenous interests of a particular dialogical context.

Cherryholmes (1988) has shown the relevance of discourse analysis and interpretation theory to the inference that connects an event with a quantity, the measured with the measuring, rejoining educational and social measurement to its social roots in a manner that deepens and broadens Gadamer's construal of conversation as an art of testing. In focusing on construct validity, as this was introduced by Cronbach and Meehl (1955), Cherryholmes has emphasized that measures are themselves interpretations, and that we are continually faced with the problem of interpreting interpretations. She (1988, pp. 449-450) follows Rorty in distinguishing between two ways of making sense out of things in this bottomless flux of texts. The first way tells stories of how the people involved have made a contribution to their community, drawing from and adding to their common history and language, exemplifying solidarity. The second way tells stories that appeal to an objective reality that exists outside of the community, and would have the ongoing conversation spoken only in the language imposed from this inhuman realm. The second way of making sense of our many-layered interpretations corresponds to the traditional reading of the book of nature as a God-given, inalterable text; in this case, instruments are one-sided and facts eternally immutable. The first way corresponds to the recognition that the facts of science are subject to historical change and that what we are willing to call factual is bound up in an intricate interplay with the myths and metaphors structuring social life; in this case, the instruments of education and science are two-sided, dialectical and dialogical.

As Cherryholmes points out, even though the second way of making sense out of life appeals to an authority above and beyond the constraints of the community, the language of that authority is nevertheless in and of the community. And this

> community is elitist, control centralized; criticism is limited to experts; the social context and historical setting of the community is not discussed; constructs (the way the community is

> conceptually organized) are not chosen on ethico-political or aesthetic grounds but in terms of 'scientific' criteria; and the discourse is thought of as nonmaterial and descriptive-explanatory (Cherryholmes 1988, p. 450).

The one-sided way in which educational measurement is performed relies on the authority of experts who control the question and answer process from outside of it, according to the traditional and positivist sense of what count as scientific criteria. On the other hand, the theory of fundamental measurement (Wright 1984, 1985) demands that the functioning of test items be constantly monitored, analyzed, accepted and rejected on the basis of explicitly ethico-political and aesthetic grounds.

The contrast between the two-sided, dialogical approach and the one-sided, monological approach to validating and justifying the interpretive act of measurement is forcefully presented by Jaeger (1987) in his Presidential Address to the National Council on Measurement in Education, wherein he comments on contemporary debates in educational measurement. Jaeger (1987, p. 8) has juxtaposed two quotes that mark the ends of the continuum along which points in the contemporary debate on educational measurement are made:

> There appears to be a fundamental difference in measurement philosophy between those on the two sides of the Rasch debate .... The difference is well characterized in the writings of Benjamin Wright (1968) and E. F. Lindquist (1953). First Wright:
>
> > Science conquers experience by finding the most succinct explanations to which experience can be forced to yield. Progress marches on the invention of simple ways to handle complicated situations. When a person tries to answer a test item the situation is potentially complicated. Many forces influence the outcome - too many to be named in a workable theory of the person's response. To arrive at a workable position, we must invent a simple conception of what we are willing to suppose happens, do our best to write items and test persons *so that their interaction is governed by this conception* and then impose its statistical consequences upon the data to see if the invention can be made useful. (1968, p. 97) [emphasis added].
>
> In contrast, Lindquist wrote:
>
> A good educational achievement test must itself define the objective measured. This means that the method of scaling an educational achievement test should not be permitted to determine the content of the test or to alter the definition of objectives

> implied in the test. From the point of view of the tester, *the definition of the objective is sacrosanct*; he has no business monkeying around with that definition. *The objective is handed down to him* by those agents of society who are responsible for decisions concerning educational objectives, and what the test constructor must do is to attempt to incorporate that definition as clearly and exactly as possible in the examination that he builds. (1953, p. 35) [emphasis added].

Despite the seemingly objectivistic tone of his language regarding the way "science conquers experience" and "progress marches," Wright is explicitly appealing to the inevitability of the way we create meaning and interpret situations by focusing our attention on what can be manageably represented. An effort must be made to formulate guesses, "what we are willing to suppose happens," and submit these in the form of questions and answers to the ruling imposed by the thing itself as it plays itself out through their interaction, and then to examine the data for signs concerning the extent to which this effort has been successful. And in opposition to the apparent objectivism of this passage, Wright has elsewhere written that "there are no natural units. There are only the arbitrary units we construct and decide to use for our counting" within a particular frame of reference (Wright and Masters 1982, p. 9). When the interactive question and answer process can be governed by a single, simple conception of a line of inquiry, data live up to the requirements of fundamental measurement, the meaning of something has been successfully negotiated, and an object of discourse has been articulated in terms of arbitrary units we can count on, such that decisions concerning student ability and item difficulty can be made with more confidence than had the data not been deliberately constructed and examined for such a structure. Not only are decisions made more confidently, but this confidence is accompanied by a more humble willingness to admit new evidence as the situation changes; experience may conquer science more frequently than vice versa, and progress may dance, leap and flow more than it marches, but this does not alter the fact that the time we share together along a stretch of life's path is marked only by the meanings that emerge from between us (Ricoeur 1984, 1985).

In contrast, Lindquist explicitly appeals to an authority on high that hands down definitions and objectives, asserting that methods can be prevented from determining content in direct opposition to everything the philosophy, history and sociology of science has to say on the matter. Such is the position also assumed by Divgi (1986), Whitely and Dawis (1974) and Wood (1978); Goldstein (1979, p. 218) is particularly adamant in his opposition to the use of Rasch and Wright's ideas on fundamental measurement, saying "that the criteria which properly ought to determine the content of an educational test are primarily educational rather than statistical," a comment like Lindquist's opening line that can be superficially construed to appear in accord with a concept of authority as emerging from within the relations studied. Goldstein and Lindquist are, however, trying to deny that methods play any role in the determination of content, as if it is possible to measure by decree, or to legislate in committee what shall count as a valid construct. Lindquist and Goldstein are correct in saying that the test must itself define what is measured, but they fail to ask just what the test is, assuming that it is what the experts say it is and not what is played out by the community of speakers participating in the language game. This is to deny that questioning inherently follows after an object always already implied in the interplay of social relations, and to assert that a line of questions can be made reliable and valid for measuring ability without entering into the circular and mutual implication of subject and object. Lindquist and Goldstein presume that fundamental measurement procedures are just another statistical technique or methodology; if this was the case, their argument might hold some water, but what makes these measurement procedures fundamental is that they are a model of how things come to a stand in the flow of experience. That nothing ever comes to an absolute and immovable halt only means that we must be more vigilant in the attention we accord the flow.

The practical import of the difference between Wright and Lindquist's positions is that educational research is too often making do with low quality data and double standards instead of stepping into the hermeneutic circle with a willingness to question the questions even more

rigorously than the respondents are questioned. Lindquist and Goldstein's position implies that measurement is more a matter of finding a mathematical description of data as they are given than it is a matter of checking and rechecking that the data can actually be counted on to provide reliable and valid comparisons. They want to fit models to data rather than the more hermeneutically sound fitting of data to measurement models because the test content has been handed down from on high with a validity that cannot be questioned by the scaling method nor by the students responding. This is apparently why two- and three-parameter models offered by Item Response Theory (IRT) are so popular and why their advocates are willing to routinely make decisions as to the applicability of a model to data on the basis of the number of items on the test that meet the assumptions of the model. In the judgment of IRT, "the one-parameter [Rasch] model has rarely provided a satisfactory fit to the test data; the three-parameter model nearly always has" (Hambleton and Rogers 1989, p. 148; also see Whitely and Dawis 1974, p. 175) because item content is allegedly sacrosanct, handed down from an unquestionable authority, and immune from influences introduced by the method of scaling. According to the IRT perspective on educational measurement, fundamental measurement theory 'assumes' so much about the data as to be virtually unworkable, but what appear to be excessively demanding assumptions in IRT turn out to be the requirements for meaningfulness delineated by both fundamental measurement theory (Suppes and Zinnes 1963, Luce and Tukey 1964, Perline, Wright and Wainer 1979, Luce and Suppes 1981, Brogden 1977) and discourse analysis (Cherryholmes 1988, Gadamer 1989, Ricoeur 1977, 1981, Weinsheimer 1985, Heelan 1983, Ihde 1979).

The IRT approach to measurement is too often willing to overlook the fact that there are commonly, if not always, substantive and interesting reasons why some test items might not point in the same direction as the others, and that these reasons are diagnostic aids in improving the functioning of the test. The authors of these studies frequently say that the data at hand does not

meet the "assumptions" of the "one-parameter model" and use this as a reason for recommending the use of another model that the data fit better; do these writers actually accept the authority of the data with no questions concerning whether children were making jokes out of their responses, or whether some of the test items might not actually pertain to the object of discourse? One would hope not, but one rarely, if ever, sees one of these authors recommend strategies for altering the test content; to do so, of course, would render the two- and three-parameter models irrelevant, but these models are already relevant only to the extent that they manage to minimize the number of prickly questions that arise from the object of interest. By including parameters that describe variations in item discrimination and in the propensity for an item to prompt guessing among the respondents, the three-parameter model closes off potentially important questions that might otherwise arise about the unity of the arrow of meaning allegedly delineated. Instead of provoking questions concerning the nature of the object to which the questions and answers are presumed to attend, the three-parameter model makes it easy to do as Lindquist prescribes, which is not to monkey around with the dictates of the authorities on high who hand down sacrosanct educational objectives.

If this is the case, though, we have to wonder just what kind of political and moral climate is engendered in education by the approaches to testing advocated by IRT. Is it really necessary, helpful, more meaningful or useful to define educational objectives according to criteria that have the source of their authority outside of the question and answer dynamic of testing? Would it be less productive and interesting to allow each local conversation to set its own standards and then study the commonalities and divergences among these? Rather than assuming discourse to be nonmaterial and descriptive-explanatory, can it be denied any longer that discourse constitutes the object of interest? Can the ethico-political and aesthetic criteria by which all constructs are chosen continue to go unexamined in educational testing? Could it be that educational measurement tends to attract people who like being part of an elite exercising centralized control

over their community, who are endowed with an unquestionable expertise and can afford to ignore the social context and historical setting of their community?

The desirable answer to each of these questions is "no," but there are many things about the current state of educational measurement that suggest the answers actually being given are "yes." The two- and three-parameter models are more difficult and more expensive to use than the models based on fundamental measurement theory, as their advocates themselves admit: they require large samples of respondents and many test items (Bejar in Hambleton 1983, p. 3; Ironson in Hambleton 1983, p. 160; and Wingersky in Hambleton 1983, p. 46; all cited by Wright 1984, p. 286), and even then not all of the parameters may be accurately estimated (Hulin, Drasgow and Parsons 1983, p. 100; cited by Wright 1984, p. 286). Furthermore, the "difficulty in applying these models is stressed," "working with [these models] is an arduous process," "the largest hurdle is the parameter estimation," the "procedures involved in parameter estimation are complex and require sophistication on the part of the user," and the potential user "must be ready to pay ... by investing substantial resources in parameter estimation and model monitoring" (Bejar in Hambleton 1983, pp. 1, 3 and 17; cited by Wright 1984, p. 286). More recently, Hambleton and Rogers (1989, p. 158) say that

> Computer programs are still not readily available, user friendly, or always going to provide the desired output. Bugs in the computer programs are an on-going problem, and computer time can be excessive. Sample sizes are sometimes prohibitive, too. No one will give you a lower bound estimate for a sample-size to jointly estimate ability and three item parameters, but it must surely be several hundred or more ....

These arduous, complex, sophisticated and expensive hurdles and processes elevate IRT specialists to an elite group that has centralized control over their sphere of interest and upon which is conferred an unquestionable scientific authority. The powers that be have traditionally used science to legitimate and justify the decisions they impose; knowledge is power, and even when knowledge is obtained from within the authoritative dynamic of local conversations it may be

surreptiously applied to limit and constrain the possibilities present for those involved. And even though it may be argued that test content is determined by examination and curriculum research, how much of this has been conducted in a hermeneutically sound fashion and how much of it is made available to the students and teachers involved?

Also, how eager can the elite of educational measurement be to introduce rigorous measurement techniques into the classroom if they continue to disdain what they themselves admit are the simplest, most elegant and parsimonious approaches in favor of the most complicated, difficult and expensive? Even those who allow for only a limited application of fundamental measurement models in their own work praise their statistical virtues (Swaminathan in Hambleton 1983, pp. 30, 33, and 35; cited by Wright 1984, p. 286), and say that they work well on all types of data (Hulin, Drasgow and Parsons 1983, pp. 57, 95, and 96; Hambleton 1983, pp. 200, 221 and 226; cited by Wright 1984, p. 286). Hambleton and Cook (1977, p. 76) say that work on virtually all psychometric models except those of Rasch and Wright is addressed to theoreticians rather than practitioners, that Rasch and Wright's work has generated considerable enthusiasm among practitioners, and that even at the time of that writing (1977) there was a "fast and convenient-to-use computer program" for applying fundamental measurement theory to real data. Today's programs (Wright and Linacre 1990, Wright, Linacre and Schultz 1990) are even faster and more convenient to use.

Hambleton and Cook (1977, p. 88) further admit that the problem of parameter estimation "is essentially resolved" in fundamental measurement theory, in contrast to the problems of parameter estimation associated with the two- and three-parameter item response theory models. Van de Vijver (1986, p. 45) concurs, saying that

> The often reported lack of convergence of the estimation procedures for small data sets [associated with many psychometric methods] does not occur in the Rasch model. Even if in more-parameter models the iterative procedure converges, not all parameters may be estimated correctly.

Through the conjoint structure of fundamental measurement, educators gain the ability to step into the circular hermeneutic of instrumentation on the same foot as the methodological imperatives issued by the emergent nature of the subject matter itself. In accord with Suppes and Zinnes' (1963) comments as to the first problem of measurement, the primary goal of all research becomes the calibration of reliable and valid instruments. Instead of trying to control for each of the many factors that can influence the construction of a data set from outside of the circularity involved in the assignment of numbers to qualities, the hermeneutic of instrumentation (Heelan 1983) embodied by conjoint measurement models offers the possibility of making "the scientific theme secure by working out [the] fore-structures [of understanding] in terms of the things themselves" (Heidegger 1962, p. 195). To be off on the right foot in the hermeneutic of instrumentation is not to blindly accept authority as it is handed down, but to question the factuality of things such that reliability and validity are determined by the extent to which questions and answers hang together along a continuum of more and less that emerges from the data itself and serves as its own indigenous ordering principle.

Two- and three-parameter IRT models can close off questions concerning the nature of the relationship between test items and the object of discourse because they "destroy the possibility of explicit invariance of the estimates of the person and the item parameters" (Andrich 1988, p. 67). These models suit themselves to propaganda purposes in the same way that any methodology can when it degrades or ignores the role of imagination in discourse, as Gadamer (1976, p. 11; emphasis added) points out:

> Statistics provide us with a useful example of how the hermeneutical dimension encompasses the entire procedure of science. It is an extreme example, but it shows us that science always stands under definite conditions of methodological abstraction and that the successes of modern sciences rest on the fact that other possibilities for questioning are concealed by abstraction. This fact comes out clearly in the case of statistics, for the anticipatory character of the questions statistics answer make it particularly suitable for propaganda purposes. Indeed, effective propaganda must always try to influence initially the judgment of the person addressed and to restrict his possibilities of judgment. Thus what is addressed by statistics seems to be a language of facts, but *which questions these facts answer and which facts would*

> *begin to speak if other questions were asked are hermeneutical questions.* Only a hermeneutical inquiry would legitimate the meaning of these facts and thus the consequences that follow from them.

Objectivity was never strictly a matter of imposing an order defined by a source outside of the context at hand, but now it is possible to account for the limits and possibilities opened up by instruments bearing the property of specific objectivity, as Rasch (1960, 1977) called it. Cherryholmes (1988) has very helpfully provided a text that draws out some of the implications of what this passage from Gadamer means for statistical construct validity in general and for educational measurement in particular, but the value of Rasch and Wright's work is that it has taken several crucial steps toward the integration of qualitative, interpretive sensibilities with quantitative, factual ones.

Conclusion. Gadamer's elaboration of the manner in which conversation is an art of testing is part of the overall goal he shares with his teachers Husserl and Heidegger: correcting the imbalance between our moral and technical understandings by restoring science's meaning for life. Husserl (1970, pp. 49, 353-378) put this project in terms of recovering Galileo's 'fateful omission,' the lack of an explicit articulation of the means by which the procedures of scientific measurement were formulated. Heidegger (1962, 1967) rethought the project, emphasizing the hermeneutic rule that factual details are always understood in terms of some larger whole, and that the whole is in turn understood in terms of its parts.

In the course of his analysis, Heidegger (1967, p. 93; 1977a, p. 269) describes the evolution of mathematical thinking in a way that displays the transition from the kind of testing that can be said to take place in everyday experience to that taking place in scientific experimentation. Though he does not explicitly refer to this passage in Heidegger, Gadamer is to be understood as engaged in the same task of demythologizing science, as he (1981, p. 150) calls it, when he makes the observation that conversation is an art of testing. However, both Heidegger and

Gadamer have often been read as rejecting a critical relation to science, and this reading raises at least two problems in the previous presentation of Gadamer's sense of conversation.

First, Gadamer can appear to reject the possibility of a methodological hermeneutics, leading Ricoeur (1981, pp. 60, 131) to suggest that *Truth and Method* may have been better titled *Truth or Method*, because Gadamer's theme is that truth always exceeds method, and insofar as anything is found to be true, the path followed always eludes methodological routinization. A closer reading of Gadamer reveals, however, that what he is rejecting is not method itself, but what has been called the 'Cartesian anxiety' (Bernstein 1983, Wachterhauser 1988, pp. 233-234) that it is impossible to formulate a set of rules that will always and everywhere produce factual information. Methodological rules that, when retraced, always lead to the same results cannot but be imposed from outside of the local, emergent interplay, and this is the sense of method Gadamer (1989, pp. 459-460, 463-5) rejects.

The sense of method he aims to describe is one that joins Hegel in re-invigorating the ancient Greek juxtaposition of following-after (*meta-*) with the road (*odos*) taken up by the thing itself; "the true method was an action of the thing itself" (Gadamer 1989, p. 464). Concepts launch this action in a playful way that draws the speakers into the language game, removing the burden of initiating the interaction and leading toward the recreative, transformative possibilities offered when people share themselves on the path of meaning. The practical import of method as the playful action of the thing itself is stressed by Ricoeur (Gadamer and Ricoeur 1982, p. 317) in his response to a question put to him by Gadamer:

> What can we do with a philosophy of dialogue if it is not able to be reconnected with the discipline of the human sciences, if it is merely a face-to-face relationship, and if it cannot provide us with, if it cannot structure, an epistemology? The risk would otherwise be that we would oppose truth to method, instead of rethinking method itself according to the requirements of truth.

Gadamer's response to this was to exclaim, "Well, I thank you for your full agreement! That was exactly what I had in mind ...." Gadamer's effort is aimed at casting doubt on the idea that what is true is only what can be methodically proven; as Weinsheimer (1985, pp. 49-52) shows, Gadamer's assertion that truth exceeds method is supported by Godel's theorem, which shows the same thing in mathematical terms. Methodical proofs are no longer the only way in which the truth of a theorem may be decided; some mathematical truths cannot be formally demonstrated, but remain formally undecidable.

The effect of this situation is re-emphasize the ancient root meaning of proof as a test of goodness, and not a elaboration of every step involved in the production of a result; this meaning is still apparent in the reference to the exception that proves the rule. Such an implication also survives in many words that have 'prove' as a root: reprove, probation, improve, probate, etc. The measurement models offered by Rasch and Wright are probabilistic in the sense of testing the goodness of the questions and answers, as well as in the sense of themselves always being "on trial," as Rasch (1960, p. 38) put it. Probabilistic measurement models not only test the goodness of what is said by checking the extent to which it sticks to the topic, but their own relevance is constantly challenged by our capacity to arrange observational frameworks that will meet their requirements. For these reasons, probabilistic models can never be the final say on what measurement is or can be. Neither do they presume to dictate a single, over-arching method of how questions must be formulated and posed, because this is a matter that can be determined only by the object of discourse itself. Insofar as method is a following-after on the path of meaning delineated by the thing itself, however, the manner by which partners to a conversation find themselves together on this path can be modeled, and this is what Rasch has accomplished. As Hirsch (1967, p. 25; also see Ricoeur 1981, p. 212-213; 1976, p. 79) puts it, "the methodological activity of interpretation commences when we begin to test and criticize our guesses." In educational measurement, examiners guess as to whether the questions they ask actually pertain

to the object of discourse and they guess as to the appropriateness of the questions for the respondents. Though the hermeneutic attitude by which guesses are formulated remains beyond the scope of methodology, insofar as method does not admit of a hermeneutic moment, we fail to check and recheck whether fancies, popular opinions and slanted questions are being offered and accepted in lieu of the thing itself.

If this first problem regarding the application of the hermeneutic attitude to quantitative measurement stems from Gadamer's hesitations concerning the possibility of finding sufficient consistency in questioning to justify the formulation of a method, the second problem stems from his critics' assertions that he over-emphasizes the unity of tradition at the expense of the discontinuities and exclusions that also characterize so much of the conversation that we are. Derrida (1978, 1981, 1982), for instance, turns away from truth to the notion of dissemination, wherein what is said is constantly breaking up and aggregating into mutually incommensurable discourses; he views hermeneutics as dangerous because it is too easily construed as a domestication of otherness, an effort that always reduces the unfamiliar and strange to what is already known. As Caputo (1987, p. 261) says, "Gadamer tends to ignore the subversion of hermeneutic *phronesis* by a diversity of power plays."

To portray hermeneutics in this way, though, is to underemphasize Gadamer's focus on play and the recreative transformation of self that occurs in the encounter with the other. Weinsheimer points out that Gadamer's reasons for taking play as the primary ontological clue coincide with Derrida's (1978, pp. 292-293) analysis of the differences between

> 'two interpretations of interpretation, of structure, of sign, of play. The one seeks to decipher, dreams of deciphering a truth or an origin which escapes play.... The other, which is no longer turned toward the origin, affirms play and tries to pass beyond man and humanism....' Excepting Derrida's accentuation of the difference between play and truth, Gadamer would find much to agree with here (Weinsheimer 1985, p. 101).

The first interpretation of interpretation corresponds to that unauthentic mode of being that appeals to an authority outside of the dialogue for its legitimation; the second is the one that is willing to let the play of signifiers have its way en route to a more homegrown and local determination of what authentic being is.

Derrida's position may not offer markedly more radical or authentic discursive strategies than Gadamer's, as may be seen in Hoy (1986), Gadamer (1986), Caputo (1987), and Dallmayr (1987, pp. 130-158). This is made apparent by Derrida's own comment that he tries "to place [him]self at a certain point at which -- and this would be the very 'content' of what I would like to 'signify' -- the thing signified is no longer easily separated from the signifier" (Derrida in Wood and Bernasconi 1988, p. 88). It may be too facile a reading of Gadamer to say that he takes the action of the thing itself to delineate the separation of signifier and signified simply and easily. If this is the case, why is "the real power of hermeneutical consciousness ... our ability to see what is questionable" (Gadamer 1976, p. 13), and why is the source of rigor in the hermeneutic experience *uninterrupted listening* (Gadamer 1989, p. 465)? Conversely, Derrida cannot communicate the difficulties of separating signified from signifier except through efforts that are aimed at facilitating that separation.

If Gadamer, in contrast to Derrida, tries to place himself at a point at which the thing signified is easily separated from the signifier, why would he stress that the questioner becomes the one questioned according to the hermeneutic logic of the question (Gadamer 1989, p. 461)? Why should he agree with Heidegger that to be guided by the things themselves is not something that can be decided once and for all, but must be our "first, last, and constant task" (Gadamer 1989, pp. 266-267) if it is not because he is also concerned with maintaining a wakeful involvement in the creation of meaning (which is often the same thing as pointing out disruptions of meaning)? And far from focusing exclusively on the search for unity of meaning in a text, Gadamer must employ a deconstructive moment in his hermeneutic if only because interpretation would self-

destruct if it was not also engaged in opening up questions on the basis of the contradictions and tensions that arise from the reading of the text. So an alternative reading of Gadamer asserts that his text is just as concerned with making the old and familiar seem new and strange as he is with finding something intimate in what is fresh and impertinent.

Educational measurement cannot be fundamental except to the extent that it playfully allows the object of discourse to represent itself through the players involved; only then will the players begin to realize themselves, be drawn up into the game and therein experience the fulfillment of re-creation. The playful loss of subjective awareness most fundamentally manifests itself in learning when the recognition of facts is most urgent, for instance in the communication or reading of signs of danger.

> We say that we 'conduct' a conversation, but the more genuine a conversation is, the less its conduct lies within the will of either partner. Thus a genuine conversation is never one that we wanted to conduct. Rather, it is generally more correct to say that we fall into conversation, or even that we become involved in it. The way one word follows another, with the conversation taking its own twists and reaching its own conclusion, may well be conducted in some way, but the people conversing are far less the leaders of it than the led. No one knows in advance what will 'come out' of a conversation (Gadamer 1989, p. 383).

In the same way, it is said that people fall into song, fall in love, fall into step and that things fall into place or into line. Carse (1986, p. 15) says that

> We are playful ... when there is no telling in advance where our relationship with them [others] will come out - when, in fact, no one has an outcome to be imposed on the relationship, apart from the decision to continue it.

In so falling into play, we participate in the self-representative activity of the things themselves that guide, focus and direct experience, thought and attention. The dialectic of participation in things is effected and drawn along or played out by the historical conditions of its context, giving it a spirit of its own that makes things not only happen to us, but also makes them a fundamental happening of what we are ourselves. In playing for the sake of recreation, we play in order to experience the transformation of self that occurs in play. The practical import of the theory of

fundamental measurement offered by Wright is to make possible a wider accessibility to this kind of experience in education, but the deployment of playfulness requires more than just methods. It requires the imagination to overcome our apparent cultural need for some form of external authority, foundation, validation and legitimation, and to rely upon the creative process from which and in which we are born to play itself out through us. Only then will we begin to address the needs of those cutoff from participation in the conversation that we are and to be confident and secure enough in our own form of life to let others live out theirs in their own ways.

References

Ackermann, John Robert. (1985), Data, Instruments, and Theory: A Dialectical Approach to Understanding Science. Princeton: Princeton University Press.

Andrich, David. (1988), Rasch Models for Measurement. Newbury Park, CA: Sage.

Bernstein, Richard J. (1983), Beyond Objectivism and Relativism: Science, Hermeneutics and Praxis. Philadelphia: University of Pennsylvania Press.

Black, Max. (1962), Models and Metaphors. Ithaca: Cornell University Press.

Boyd, Richard. (1979), "Metaphor and theory change: What is 'metaphor' a metaphor for?", in Andrew Ortony (ed.), Metaphor and Thought. Cambridge: Cambridge University Press, pp. 356-408.

Brogden, H. E. (1977), "The Rasch model, the law of comparative judgment and additive conjoint measurement", Psychometrika 42, 631-634.

Caputo, John D. (1987), Radical Hermeneutics: Repetition, Deconstruction and the Hermeneutic Project. Bloomington, IN: Indiana University Press.

Carraher, Terezinha Nunes. (1989), "Negotiating the results of mathematical computations", International Journal of Educational Research 13(6), 637-646.

Carse, James P. (1986), Finite and Infinite Games: A Vision of Life as Play and Possibility. New York: Free Press.

Cherryholmes, Cleo. (1988), "Construct validity and the discourses of research", American Journal of Education 96(3), 421-457.

Comaroff, Jean. (1985), Body of Power, Spirit of Resistance. Chicago: University of Chicago Press.

Cronbach, Lee and Paul Meehl. (1955), "Construct Validity in Psychological Tests", Psychological Bulletin 52(4), 281-302.

Dallmayr, Fred R. (1987), Critical Encounters: Between Philosophy and Politics. Notre Dame, IN: University of Notre Dame Press.

Derrida, Jacques. (1978), Writing and Difference. Alan Bass (trans.). Chicago: University of Chicago Press.

_____. (1981), Dissemination. Barbara Johnson (trans.). Chicago: University of Chicago Press.

_____. (1982), Margins of Philosophy. Alan Bass (trans.). Chicago: University of Chicago Press.

Descartes, Rene. (1971), Philosophical Writings. Elizabeth Anscombe and Peter Thomas Geach (eds. and trans.), London: Nelson, 1954; reprint, Indianapolis: Bobbs-Merrill.

Divgi, D. R. (1986), "Does the Rasch model really work for multiple choice items? Not if you look closely", Journal of Educational Measurement 23(4), 283-296.

Duncan, Otis Dudley. (1984), Notes on Social Measurement: Historical and Critical. New York: Russell Sage Foundation.

Finn, Jeremy D. (1989), "Withdrawing from school", Review of Educational Research 59(2), 117-142.

Gadamer, Hans-Georg. (1975), "Hermeneutics and social science", Cultural Hermeneutics 2, 307-316.

_____. (1976), Philosophical Hermeneutics. David E. Linge (ed. and trans.). Berkeley: University of California Press.

_____. (1981), Reason in the Age of Science. Frederick G. Lawrence (trans.), Studies in Contemporary German Social Thought. Thomas McCarthy (ed.), no. 2. Cambridge: MIT Press.

_____. (1986), "Text and interpretation", in Hermeneutics and Modern Philosophy. Dennis J. Schmidt. (trans.). Brice R. Wachterhauser (ed.). New York: State University of New York Press, pp. 377-396.

_____. (1989), Truth and Method. 2d revised edition. Translation revised by Joel Weinsheimer and Donald Marshall. New York: Crossroad Publishing.

Gadamer, Hans-Georg and Paul Ricoeur. (1982), "The conflict of interpretations", in Phenomenology: Dialogues and Bridges. Ronald Bruzina and Bruce Wilshire (eds.). Albany: State University of New York Press.

Gerhart, Mary and Allan Russell. (1984), Metaphoric Process: The Creation of Scientific and Religious Understanding. Foreword by Paul Ricoeur. Fort Worth: Texas Christian University Press.

Goldstein, Harvey. (1979). "Consequences of using the Rasch model for educational assessment", British Educational Research Journal 5(2), 211-220.

Hacking, Ian. (1983), Representing and Intervening: Introductory Topics in the Philosophy of Natural Science. Cambridge: Cambridge University Press.

Hambleton, Ronald, Ed. (1983), Applications of Item Response Theory. Vancouver, BC: Educational Research Institute of British Columbia.

Hambleton, Ronald and Linda L. Cook. (1977), "Latent trait models and their use in the analysis of educational test data", Journal of Educational Measurement 14(2): 75-96.

Hambleton, Ronald K. and H. Jane Rogers. (1989), "Solving criterion-referenced measurement problems with item response models", International Journal of Educational Research 13(2), 145-160.

Heelan, Patrick. (1983a), "Natural science as a hermeneutic of instrumentation", Philosophy of Science 50(June), 181-204.

_____. (1985), "Interpretation in physics: Observation and measurement", Greater Philadelphia Philosophy Consortium, March.

_____. (1989), "Hermeneutics and natural sciences: Problems and prospects--Commentary", Paper read at the American Philosophical Association (Central) Meeting, Chicago, April.

Heidegger, Martin. (1962), Being and Time. John Macquarrie and Edward Robinson (trans.), New York: Harper & Row.

_____. (1967), What is a Thing? W. B. Barton, Jr. and Vera Deutsch (trans.), analytic afterword by Eugene Gendlin. South Bend, IN: Regnery.

_____. (1977a), Basic Writings. Ed. David F. Krell. New York: Harper & Row.

_____. (1977b), The Question Concerning Technology and Other Essays. Trans. William Lovitt. New York: Harper & Row.

_____. (1982), Nietzsche. Vol. 4: Nihilism. Ed., with notes and analysis, David Farrell Krell. Trans. Frank A. Capuzzi. San Francisco: Harper & Row.

Hesse, Mary. (1970), Models and Analogies in Science. Notre Dame: University of Notre Dame Press.

Hirsch, E. D. (1967), Validity in Interpretation. New Haven: Yale University Press.

Holton, Gerald. (1988), Thematic Origins of Scientific Thought. Revised ed. Cambridge: Harvard University Press.

Hoy, David Couzens. (1986), "Must we say what we mean? The grammatological critique of hermeneutics", in Hermeneutics and Modern Philosophy. Brice R. Wachterhauser (ed.). Albany, NY: State University of New York Press, pp. 397-415.

Hulin, C. L., F. Drasgow and C. K. Parsons. (1983), Item Response Theory. Homewood, IL: Dow Jones-Irwin.

Husserl, Edmund. (1970), The Crisis of European Sciences and Transcendental Phenomenology. Trans. David Carr. Evanston: Northwestern University Press.

Ihde, Don. (1979), Technics and Praxis. Boston: D. Reidel.

Jaeger, Richard M. (1987), "Two decades of revolution in educational measurement!?", Educational Measurement: Issues and Practice 6(2), 6-14.

Kuhn, Thomas S. (1961), "The function of measurement in modern physical science", Isis 52(168), 161-193.

_____. (1970), The Structure of Scientific Revolutions. 2d ed. Chicago: University of Chicago Press.

_____. (1979), "Metaphor in science", in Andrew Ortony (ed.), Metaphor and Thought. Cambridge: Cambridge University Press, pp. 409-419.

Latour, Bruno and Steve Woolgar. (1979), Laboratory Life: The Social Construction of Scientific Facts. Beverly Hills: Sage.

Lindquist, E. F. (1953), "Selecting appropriate score scales for tests (Discussion)", Proceedings of the 1952 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service.

Livingston, Eric. (1986), The Ethnomethodological Foundations of Mathematics. London: Routledge & Kegan Paul.

Luce, R. Duncan and Patrick Suppes. (1981), "Theory of Measurement", S.V. in The New Encyclopedia Britannica. 15th Edition. Chicago: Encyclopedia Britannica.

Luce, R. Duncan and John W. Tukey. (1964), "Simultaneous conjoint measurement: A new kind of fundamental measurement", Journal of Mathematical Psychology 1(1), 1-27.

Ormiston, Gayle and Raphael Sassower. (1989), Narrative Experiments: The Discursive Authority of Science and Technology. Minneapolis: University of Minnesota Press.

Perline, Richard, Benjamin D. Wright and Howard Wainer. (1979), "The Rasch model as additive conjoint measurement", Applied Psychological Measurement 3(2), 237-255.

Perret-Clermont, Anne-Nelly and Maria-Luisa Schubauer-Leoni. (1989), "Social factors in learning and teaching: Towards an integrative perspective", International Journal of Educational Research 13(6), 575-580.

Rasch, Georg. (1960), Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danmarks Paedogogiske Institut; reprinted, with Foreword and Afterword by Benjamin D. Wright, Chicago: University of Chicago Press, 1980.

_____. (1964), "Objective comparisons", Lecture given at the UNESCO seminar, Voksenasen, Oslo, Norway. Mimeographed.

_____. (1977), "On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements", Danish Yearbook of Philosophy 14, 58-94.

Ricoeur, Paul. (1976), Interpretation Theory: Discourse and the Surplus of Meaning. Fort Worth: Texas Christian University Press.

_____. (1977), The Rule of Metaphor: Multi-Disciplinary Studies of the Creation of Meaning in Language. Trans. Robert Czerny. Toronto: University of Toronto Press.

_____. (1981), Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation. Ed., trans. and intro. by John B. Thompson, with a response from the author. Cambridge: Cambridge University Press.

_____. (1984), Time and Narrative. Vol. 1. Trans. Kathleen McLaughlin and David Pellauer. Chicago: University of Chicago Press.

_____. (1985), Time and Narrative. Vol. 2. Trans. Kathleen McLaughlin and David Pellauer. Chicago: University of Chicago Press.

Rothbart, Daniel. (1984), "The semantics of metaphor and the structure of science", Philosophy of Science 51, 595-615.

Schubauer-Leoni, Maria-Luisa, N. Bell, M. Grossen, and Anne-Nelly Perret-Clermont. (1989), "Problems in assessment of learning: The social construction of questions and answers in the scholastic context", International Journal of Educational Research 13(6), 671-684.

Suppes, Patrick and Joseph L. Zinnes. (1963), "Basic measurement theory", in R. Duncan Luce, Robert R. Bush and Eugene Galanter (eds.), Handbook of Mathematical Psychology. New York: John Wiley & Sons, pp. 1-111.

Toulmin, Stephen. (1982), "The construal of reality: Criticism in modern and postmodern science", Critical Inquiry 9(September): 93-111.

Van de Vijver, Fons J. R. (1986), "The robustness of Rasch estimates", Applied Psychological Measurement 10(1): 45-57.

Voigt, Jorg. (1989), "Social functions of routines and consequences for subject matter learning", International Journal of Educational Research 13(6), 647-656.

Weinsheimer, Joel C. (1985), Gadamer's Hermeneutics: A Reading of "Truth and Method." New Haven: Yale University Press.

Whitely, Susan E. and Rene V. Dawis. (1974), "The nature of objectivity with the Rasch model", Journal of Educational Measurement 11(2): 163-178.

Wood, David and Robert Bernasconi (eds.). (1988), Derrida and *Differance*. Evanston: Northwestern University Press.

Wood, Robert. (1978), "Fitting the Rasch model: A heady tale", British Journal of Mathematical and Statistical Psychology 31, 27-32.

Woods, Peter. (1989), "Opportunities to learn and teach: An interdisciplinary model", International Journal of Educational Research 13(6), 597-606.

Wright, Benjamin D. (1968), "Sample-free test calibration and person measurement", in Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, pp. 85-101.

_____. (1977), "Solving measurement problems with the Rasch model", Journal of Educational Measurement 14(2), 97-116.

_____. (1984), "Despair and hope for educational measurement", Contemporary Education Review 3(1), 281-288.

_____. (1985), "Additivity in psychological measurement", in Edward Roskam (ed.), Measurement and Personality Assessment. North Holland: Elsevier Science Publishers, pp. 101-111.

Wright, Benjamin D. and John M. Linacre. (1990), FACETS: A Computer Program for Many-Faceted Rasch Analysis. Chicago: MESA Press.

Wright, Benjamin D., John M. Linacre and Matthew Schultz. (1990), BIGSCALE: A Rasch-Model Rating Scale Analysis Computer Program. Chicago: MESA Press.

Wright, Benjamin D. and Geofferey Masters. (1982), Rating Scale Analysis. Chicago: MESA Press.